A "Quasi-Flexible" Automatic Docking Processing for Studying Stereoselective Recognition Mechanisms. Part I. Protocol Validation

S. ALCARO,[1] F. GASPARRINI,[2] O. INCANI,[3] S. MECUCCI,[2] D. MISITI,[2] M. PIERINI,[2,*] C. VILLANI[2,*]

[1]*Dipartimento di Scienze Farmaco-Biologiche, Università degli Studi di Catanzaro "Magna Græcia," "Complesso Ninì Barbieri," Roccelletta di Borgia, 88021 Catanzaro, Italy*
[2]*Dipartimento di Studi di Chimica e Tecnologia delle Sostanze Biologicamente Attive, Universitá "La Sapienza," P.le A. Moro 5, 00185 Roma, Italy*
[3]*Laboratorio di Chimica Computazionale, Tecnofarmaci S.C.p.A., Via del Mare 87, 00040 Pomezia, Italy*

Received 22 July 1999; accepted 30 December 1999

ABSTRACT: The main purpose of this work is the development and validation of a general scheme based on a systematic and automatic "quasi-flexible" docking approach for studying stereoselective recognition mechanisms. To achieve our goals we explore the conformational and configurational space for small- or medium-size flexible molecules in a systematic way, seeking a method that is both reasonably accurate and relatively fast from the computational point of view. In particular, we have developed a general computational protocol for the global molecular interaction evaluation ("Glob-MolInE") to efficiently explore the orientational and conformational space of flexible selectors and selectands used in modern chiral high-performance liquid chromatography (HPLC); the enantioselective binding of the selector (*S*)-N-(3,5-dinitrobenzoyl)-leucine-*n*-propylamide (*S*)-**1** towards the selectand N-(2-naphthyl)-alanine methyl ester **2** has been studied; the global minimum obtained for the homochiral associate [*S*(**1**)/*S*(**2**)] (Pop. >99%) is very close (RMS $\simeq$ 0.20) to the crystallographically determined structure. J Comput Chem 21: 515–530, 2000

Correspondence to: F. Gasparrini; e-mail: gasparrini@uniroma1.it

*Present adress: Dipartimento di Scienze del Farmaco, Universitá "G. D'Annunzio," Via dei Vestini 31, 66013 Chieti, Italy

Contract/grant sponsors: Ministero dell'Università e della Ricerca Scientifica (MURST, Italy) and Consiglio Nazionale delle Ricerche (CNR, Italy)

Keywords: "quasi-flexible" automatic docking; molecular recognition; stereoselective discrimination; enantioselective "high-performance liquid chromatography" (HPLC); systematic configurational search

Introduction

Enantioselective chromatography is one of the major techniques for the direct resolution of racemates.[1] Modern synthetic chiral stationary phases (CSPs) are frequently designed for the resolution of certain classes of enantiomers, and several of them have been discovered by a process of trial and error. However, many others have been synthesized using the reciprocity concept[2] by which a molecule that is found to be resolvable on an existing CSP is itself attached to a solid support and tested as a CSP; several generations of such chiral sorbents have been developed by Pirkle and coworkers.[1j, 2]

Today the design and synthesis of new selectors has become an important and rapidly growing field of chemistry. The essence of any enantioselective chromatographic process depends critically on the stereoselective molecular interactions, and molecular recognition studies are concerned with such phenomena wherein elucidation of the rules and restrictions governing these intermolecular interactions is of paramount importance for the understanding and manipulation of these processes.

To have a good understanding of the chiral recognition mechanism at the molecular level, we have begun a series of very extensive computational studies. Our primary goal is to elucidate the enantioselective separation mechanism, i.e., to "understand" the binding of the given enantiomers to the known CSP in terms of intermolecular interactions. To accomplish this, we must be able to reproduce the observed thermodynamic parameter of the separation to a sufficient accuracy (0.5–1.0 kcal/mol). In a second step, using the gained "understanding" at the molecular level, predictions are made about the effect of structural variations between the interacting partners on the recognition process. From the computational point of view this becomes difficult because one has to treat the "docking problem"[3] between medium-large sized molecules, which often contain π-systems, and a number of torsional degrees of freedom.

Several chiral recognition models relating to the mode of operation of these CSPs have been proposed starting from a body of experimental data,[1, 2] and several such hypotheses have attracted the attention of people involved in the elucidation of the origin of chiral recognition in those systems by computational methods.[4–10]

Although it is clear "what" should be done in the computational approach to the CSP design it is unclear "how" this should be done in a rational and standardized way.

Therefore, the main purpose of this work is the development and validation of a general protocol based on a systematic automatic "quasi-flexible" docking approach that we call the Global Molecular Interaction Evaluation ("Glob-MolInE"). To achieve our goals we must explore the conformational and the configurational[11] space for small or medium size molecules, seeking a method that is both reasonably accurate and is relatively fast from the computational point of view.

The automatic "quasi-flexible" docking methodology developed in our laboratories can be defined as a systematic configurational searching approach using first a local rigid optimization, but followed with a final full minimization (totally relaxed molecules) on the energy minima located from the rigid optimizations. The docking procedure is performed over all the relevant conformations (normally within 3RT) of the isolated molecules (selector and selectand): we term this approach "quasi-flexible" docking.

Theory and Computational Methods

GENERAL STRATEGY

Enantiodiscrimination systems are studied according to a general strategy (Fig. 1) based on a combination of molecular mechanics techniques: conformational searching and analysis (Phase A), automatic systematic rigid body docking (Phase B) and "quasi-flexible" docking processing (Phase C).

CONFORMATIONAL SEARCHING AND ANALYSIS (PHASE A)

The problem of how to determine the conformation of flexible molecules in solution has challenged researchers for decades, and will likely continue to do so.[12] Computational chemistry, especially when

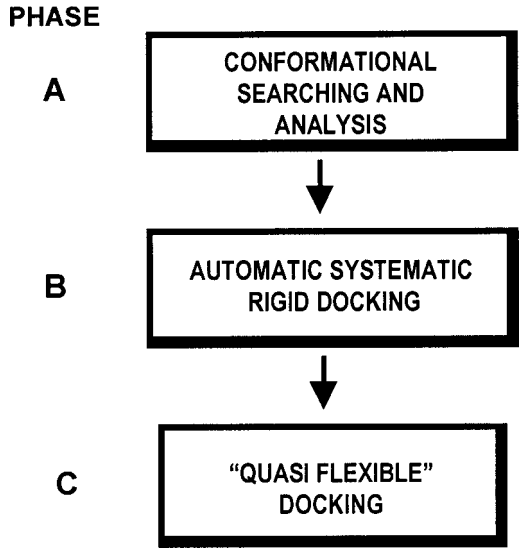


FIGURE 1. Strategy used for modeling enantioseparation.

used in combination with spectroscopic techniques, has facilitated considerable progress in conformational searching and analysis during the past 2 decades. The automatic conformational search for the selector and selectand can be performed using systematic, stochastic, or molecular dynamics techniques. These methodologies are available within commercial software using a variety of Empirical Force Fields (E.F.F.). In our particular case, we have chosen the stochastic approach.

Furthermore, the conformational analysis is essential for extremely flexible molecules, featuring many different torsional degrees of freedom. In such cases the number of different conformers found, within a 3RT window, is so large that a complete, multiconformational docking process becomes prohibitive: thus, the conformations are "filtered" according to energy and geometry based criteria, that remove low populated or closely related conformations.

AUTOMATIC SYSTEMATIC RIGID DOCKING PROCESSING (PHASE B)

The general scheme of the automatic "quasi-flexible"docking processing is reported in Figure 2. The process takes place in two main phases (B–C): in the first phase, involving the rigid docking approach, three main steps (B_I–B_{III}) are used, followed by the flexible docking (phase C).

In the first step (B_I), a systematic searching of supramolecular associates is performed over the six degrees of freedom defining both the relative positions and orientations of the molecules. This is repeated for each of the most populated conformers obtained from the previous phase (A).

At this level, a number of supramolecular associates are generated by the *L_GRID_6FD* calculation procedure (see later "Glob-MolInE" module). In other words, the selectand is rolled, in a discrete manner, on the van der Waals surface of the selector, and the interaction energy evaluated at each point. The systematic rigid-docking sampling is coupled with an algorithm (*SEL_5FD*) for the selection (step B_{II}) of the most stable representative configurations at the generated intermolecular potential energy, based on energy criteria. In the third step (B_{III}), an optimization procedure (*OPT_6FD*) finds the optimized local minimum energy configuration starting from each low-energy binary complex obtained by the selecting module *SEL_5FD*. The B_I–B_{III} steps consider nonbonding intermolecular contributions only, while treating the interacting molecules as rigid bodies. In this way, a set of configurations is obtained sorted by energy or statistical weight.

"QUASI-FLEXIBLE" DOCKING (PHASE C)

Finally (Phase C), the rigidly optimized conformations are submitted to a full minimization (totally relaxed molecules) in conjunction with a simplex routine for the optimization of the relative orientations in an iterative fashion; this procedure is carried out by means of the *RD_FULL MIN* module.

Statistical thermodynamics calculations on the computed configuration ensemble, before and after the full minimization procedure (Phase C), allow us to derive all the macroscopic thermodynamic quantities ($\Delta\Delta G$, $\Delta\Delta H$, and $\Delta\Delta S$) related to the enantioselective recognition mechanism.

The appellation of "quasi-flexible" docking derives from the combination of three fundamental characteristics: multiconformational analysis, rigid docking and full relaxation, with the latter two carried out in an iterative fashion, which allows to evaluate the final geometry of the most stable configurations and also the induced fits in flexible molecules during the complexation.

Global Molecular Interaction Evaluation ("Glob-MolInE") Modules

The computational protocol comprises four main modules: *L_GRID_6FD*, *SEL_5-FD*, *OPT_6FD*, and *RD_FULL MIN* (Fig. 2); an additional module

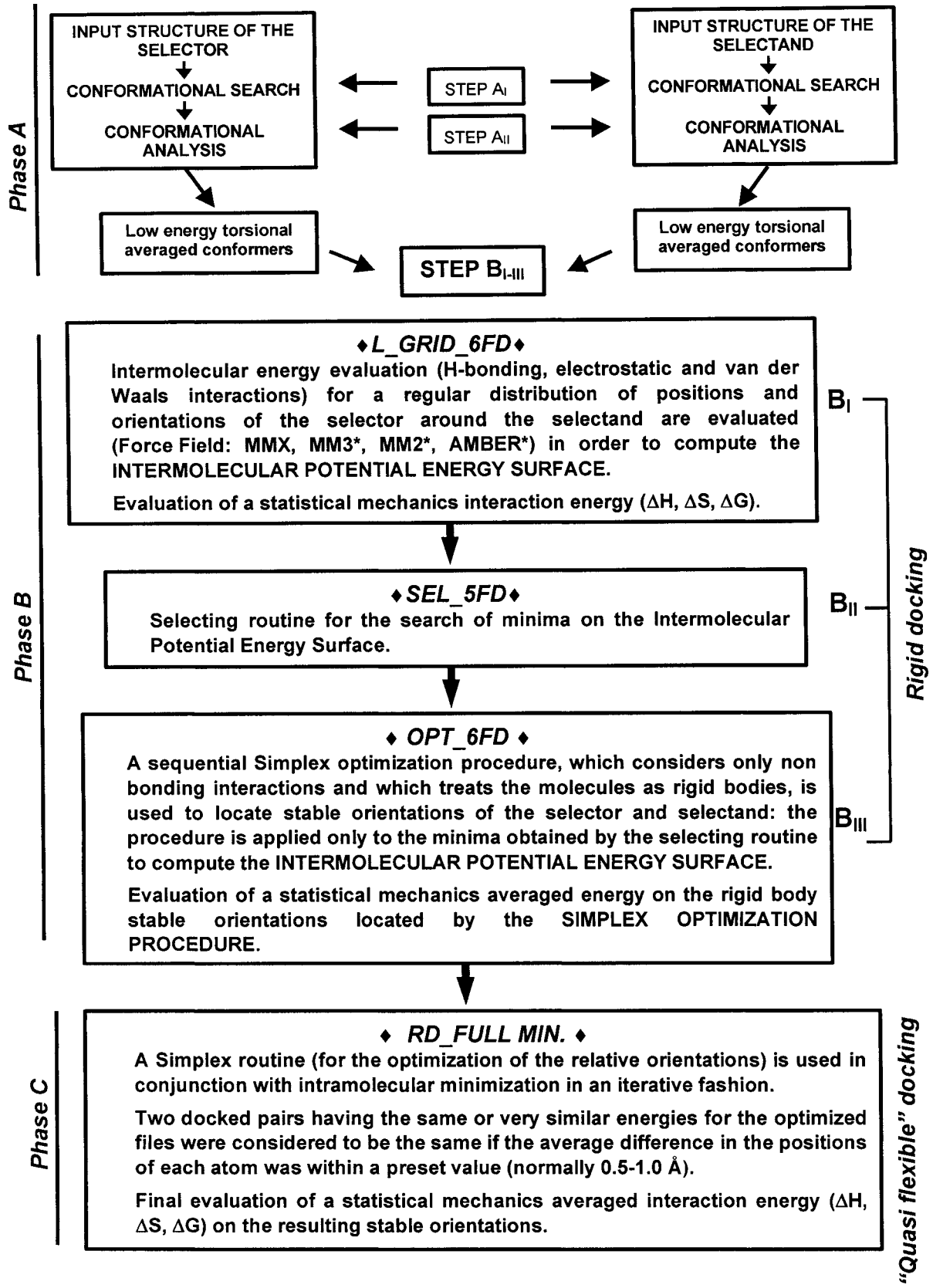


FIGURE 2. "Glob-MolInE," flow chart (an automatic "quasi-flexible" docking processing).

GM_MTE, which computes of the macroscopic thermodynamic quantities of the enantiodiscrimination process, has been used.

CONFIGURATIONS GENERATION MODULE (*L_GRID_6FD*)

The aim of the first module (*L_GRID_6FD*) is to evaluate the molecular interaction energy when the selectand is rolled, in a discrete manner, with its van der Waals surface on the van der Waals surface of the selector. The calculation is extended to all possible combinations between the selected conformers of the selector and the selectand, both considered as rigid bodies (Rigid docking).

According to a methodology similar to that already reported by Lipkowitz et al.,[4d] the *L_GRID_6FD* module performs the searching of configurations (supramolecular associates) by using a systematic procedure. In this procedure, the position and orientation of the selectand relative to the selector are assigned as follows: spherical coordinates (λ, θ, ϕ) concern the location of the center of mass of the selectand with respect to that of the selector fixed in the reference coordinate system; Euler angles (α, β, γ) specify the orientation of the selectand with respect to the selector.

The computational procedure generates an ensemble of microstates corresponding to supramolecolar associates (configurations). This is made by combining all the directions of approach (specified by θ and ϕ angles) of the selectand towards the selector with all the orientations (specified by the three Euler angles), while the sixth degree of freedom (λ, the intermolecular distance between the molecular center of mass) is fixed at the value corresponding to contact of the van der Waals surfaces of both molecules. To determine the value of λ, a simple parabolic equation is analytically solved, and the largest value between the two possible solutions is chosen as described later in details.

The directions of approach are generated in such a way to obtain a uniform spherical distribution around the center of mass of the selector. The sampling over the (θ, ϕ) polar angles is performed in such a way that the unitary vectors, determining the directions of approach, identified by these angles, are uniformly distributed on a sphere. To get a constant area for each point on this sphere, the step over θ is fixed by constant $\cos(\theta)$ increments, while the step over ϕ is equally spaced. If the translational Grid resolution (G_t) is the number of θ values (between 0 and π) and $2G_t$ the number of ϕ values (between 0 and 2π), the total number of points (T) on the sphere (i.e., the directions of approach-translational matrices) are:

$$T = G_t \cdot 2G_t + 2 = 2G_t^2 + 2 = 2 \cdot \left(G_t^2 + 1\right) \quad (1)$$

where the two "poles" are ($\theta = 0, \theta = \pi$).

In the same way, if the orientation grid resolution (G_R) is the number of β values (between 0 and π), and $2G_R$ the number of α and γ values (between 0 and 2π), the total number of relative orientations (R) is:

$$\begin{aligned} R &= 2G_R \cdot G_R \cdot 2G_R + 2 \cdot 2G_R = 4G_R^3 + 4G_R \\ &= 4 \cdot \left(G_R^3 + G_R\right) = 4G_R \cdot \left(G_R^2 + 1\right) \end{aligned} \quad (2)$$

where $2 \cdot 2G_R$ are added to take into account the condition $\beta = 0, \beta = \pi$.

The total number of supramolecular associates P is given by the following final expression:

$$P = T \cdot R = \left[2 \cdot \left(G_t^2 + 1\right)\right] \cdot \left[4G_R \cdot \left(G_R^2 + 1\right)\right] \quad (3)$$

and the accuracy of the docking processing is correlated to the resolution G_t and G_R.

Even with the rigid-body approximation, the number of possible configurations increases rapidly when increasing the resolution G_t and G_R (the computational problem belongs to the class known as nondeterministic polynomial time "NP complete" problems). Therefore, we are forced to use a moderate resolution method to contain the calculation within an acceptable CPU time.

The entity of the translation (λ) to which we submit the selectand molecule (initially at the origin of the polar coordinates) along a certain direction (θ, ϕ) is set in such a way that only a contact is established between the selector and selectand van der Waals surfaces, avoiding any atomic interpenetration. In other words, the translation of the selector is performed according to the following conditions:

$$d_{ij}(\lambda) \geq R_{ij} \qquad \forall (i, j) \quad (4)$$

with at least for one couple (i, j) the equality

$$d_{ij}(\lambda) = R_{ij} \quad (5)$$

holds true, where i [j] represents the generic atom of the selector [selectand], $d_{ij}(\lambda)$ the interatomic distance and R_{ij} the sum of van der Waals radii, while λ is the value of the distance between the mass centers that satisfies eqs. (4) and (5). In addition, the sum of van der Waals radii, R_{ij} of each ij couple of interacting atoms can be also modulated by a multiplying scale factor χ (vdW compression factor).

The interaction energies are computed using Molecular Mechanics equations. The intermolecular potential energy ΔE_i is defined as a sum of three

components:

$$\Delta E_i = \Delta E_{i(\mathrm{vdW})} + \Delta E_{i(\mathrm{Hb})} + \Delta E_{i(\mathrm{elect.})} \qquad (6)$$

where $\Delta E_{i(\mathrm{vdW})}$ is the van der Waals energy $\Delta E_{i(\mathrm{Hb})}$ is the hydrogen bond energy, and $\Delta E_{i(\mathrm{elect.})}$ is the electrostatic term, all computed pairwise additive using only two-body interaction terms.

SELECTION MODULE (*SEL_5FD*)

The number of microstates corresponding to the diastereomeric associates generated by means of *L_GRID_6FD* is always large, and requires an additional algorithm to reduce the number of configurations so that only the most representative ones are selected and submitted to the next optimization process (*OPT_6FD*). For every adduct we try to establish the position taken on the hypersurface of potential energy of interaction, which depends to the five degrees of freedom (5FD) $(\theta, \phi, \alpha, \beta, \gamma)$. On this hypersurface around every configuration (node) there are 10 other configurations nearby (first neighbours), which differ only in one of the five variables $(\theta, \phi, \alpha, \beta, \gamma)$; in other words the difference consists in a unitary increase (positive or negative) of the variables themselves. Consequently, all the configurations make up an array of nodes in a hyperspace of five dimensions in which the existing correlation among the various nodes is totally defined. In our approach, energy-minima configurations between each node and its neighbours are first identified, allowing us to select those situated at the bottom of the energetic wells. These configurations are then transmitted to the optimization process by the (*OPT_6FD*) module.

SIMPLEX OPTIMIZATION MODULE (*OPT_6FD*)

OPT_6FD is a module for the automatic optimization of the intermolecular interaction energy between the various conformations of the selector with those of the selectand. It is applied to the "configurations" obtained by the previously described procedure using the *L_GRID_6FD* module followed by the *SEL_5FD* routine. The *OPT_6FD* algorithm allows one to easily obtain rigid optimized configurations starting from those situated in the vicinity of a local energy minimum previously obtained, without the need of high resolution or long calculation times.

The energy optimization, based on the Downhill Simplex,[13] is carried out in the six-dimensional space defined by the degrees of freedom $\theta, \phi, \lambda, \alpha, \beta, \gamma$. In this step, the conformers are considered as rigid bodies. In particular, the selector is rigidly held fixed in the space, while the coordinates of the mass center (x, y, z) and the three Euler angles (α, β, γ) of the selectand, under the Simplex procedure control, are modified until the intermolecular interaction energy converges to a local minimum.

The speed and efficiency of the optimization process is conditioned by the choice of the orientation parameters, i.e., translational and rotational step values, as controlled by the Simplex algorithm during the iterations. To guarantee the location of a true local minimum, the Simplex procedure is automatically and sequentially restarted several times until a complete stability in the intermolecular interaction energy value is obtained. Using a different set of input orientation parameters in every restarting of the Simplex procedure, the intermolecular potential energy surface can be explored in detail in the proximity of every local minimum.

At the end of every optimization cycle, a selection procedure for the complete set of stable orientations can be activated to eliminate any duplicates. This procedure is based on both energy and atomic coordinates comparison. Configurations are selected for deletion if both the energy differences and the atomic RMS difference between the compared configurations is below a given, selectable tolerance value.

FULL RELAXATION MODULE (*RD_FULL MIN*)

The last module (*RD_FULL MIN*) performs an alternate rigid docking and full minimization to relax all intra- and intermolecular degrees of freedom of the system. It also eliminates duplicate structures (if any), and evaluates the energy of the relaxed diastereomeric associates.

STATISTICAL THERMODYNAMIC EVALUATION OF FREE ENERGY OF INTERACTION (*GM_MTE*)

In the enantioselective separation process the following equilibria occur throughout a chromatographic column:

$$\mathrm{CSP}^R + \mathrm{A}^R \rightleftharpoons \mathrm{CSP}^R \cdot \mathrm{A}^R \qquad [\Delta G_{RR}] \qquad (7)$$

$$\mathrm{CSP}^R + \mathrm{A}^S \rightleftharpoons \mathrm{CSP}^R \cdot \mathrm{A}^S \qquad [\Delta G_{RS}] \qquad (8)$$

where CSP = Chiral Stationary Phase, A = Chiral Analyte, and *R* and *S* are the Cahn–Ingold–Prelog descriptors.

Each equilibrium is characterized by a free energy of binding (ΔG), and $\Delta G_{R,R}$ and $\Delta G_{R,S}$ represent the free energy variation relative to the association equilibrium with the chiral stationary phase

(CSP^R) for the enantiomeric analytes A^R and A^S, respectively.

The determination of the ΔG value for equilibria (7) and (8) is of fundamental importance, because it contains all the information needed to predict the elution order of chiral analytes (the analyte more retained into the column has a more negative ΔG).

It will be assumed that the enantioselectivity factor (α) is related to $\Delta\Delta G$ by the following relation:

$$\Delta_{R,S}\Delta G = \Delta G_{R,R} - \Delta G_{R,S} = -RT \ln \alpha \qquad (9)$$

It is not necessary, however, to determine the ΔG value for each equilibrium; rather, we need to evaluate only G for the complex $CSP^R \cdot A^R$ and compare it with the G of the $CSP^R \cdot A^S$ complex; thus, eq. (9) is simplified to:

$$\Delta_{R,S}\Delta G = G_{CSP^R \cdot A^R} - G_{CSP^R \cdot A^S} \qquad (10)$$

This direct comparison is possible because the left-hand sides of both equilibria (7) and (8) are chemically identical from an energetic point of view (this is ensured by the enantiomeric relationship between the analytes), so in the differences the contribution of the free energy of the isolated selector and selectand is cancelled.

In a first approximation it is possible to divide the free energy of the complex into various contributions:

$$G_{comp.} = G_{rot.\text{-}transl.} + G_{elect.} + G_{vibr.} + G_{ster.} + G_{solv.} \qquad (11)$$

where $G_{rot.\text{-}transl.}$ = rotational-translational free energy, $G_{elect.}$ = electronic free energy, $G_{vibr.}$ = vibrational free energy, $G_{ster.}$ = steric free energy, and $G_{solv.}$ = solvation free energy.

Assuming that: (1) the surface effect of the silica microparticles can be neglected; (2) there is a 1:1 complex of CSP and the analyte; (3) upon complexation on CSP, rotational-translational degrees of freedom are completely lost; (4) the electronic and vibrational contributions are not significantly different in both the two diastereomeric complexes, and (5) solvation effects are absent or in some cases comparable for the two complexes in solution, relation (10) results in a further simplified form:

$$\Delta_{R,S}\Delta G = G_{CSP^R \cdot A^R(ster.)} - G_{CSP^R \cdot A^S(ster.)} \qquad (12)$$

The Gibbs free energy of a specific complex is a macroscopic thermodynamic quantity that represents a weighted average of all possible microscopic diastereomeric complexes; the enthalpic component (H) of the free energy is defined as:

$$H = U + PV \qquad (13)$$

where U is the internal energy of the system, and PV is the work function.

Bearing in mind that in the process under discussion no significant expansion or compression takes place, the following similarities exist: $H = U$ and $G = A$ (A = Helmholtz free energy) because $\Delta(PV) = 0$. Furthermore the calculation of $G_{ster.}$ is followed utilizing the well-known mechanical statistical relation:

$$\overline{G} = -RT \ln q \qquad (14)$$

where $\overline{G}$ = averaged free energy, R = universal gas constant, and T = temperature (K) in which q is the molecular partition function defined by the following expression:

$$q = \sum_{i}^{nA} \sum_{j}^{nB} \sum_{k}^{nO_{ij}} e^{-E_{ijk}/RT} \qquad (15)$$

having indicated with nA and nB, respectively, the number of selector and selectand conformations. With nO_{ij} we indicate the number of orientations with minimal energy obtained in the docking of the ith conformer of the selector with the jth conformer of the selectand, and with E_{ijk} we express the calculated steric energy of a microstate associated with the kth configuration of the complex between the ith conformation of the selector and with the jth conformation of the selectand. For our purposes, a microstate is any unique orientation of the selector and the selectand in a diastereomeric complex (configuration), and the value of E_{ijk} can be obtained from molecular mechanics calculations.

The averaged enthalpy and entropy are also calculated using relations derived from statistical mechanical calculations:

$$\overline{H} = \sum_{i}^{nA} \sum_{j}^{nB} \sum_{k}^{nO_{ij}} E_{ijk} \cdot P_{ijk} \qquad (16)$$

$$\overline{S} = -R \sum_{i}^{nA} \sum_{j}^{nB} \sum_{k}^{nO_{ij}} P_{ijk} \cdot \ln P_{ijk} \qquad (17)$$

P_{ijk} represents the probability of the kth orientation of the complex obtained from the pair of conformations i-j of the selector and selectand, respectively, and this is defined by the following equation:

$$P_{ijk} = \frac{e^{-E_{ijk}/RT}}{q} \qquad (18)$$

COMPUTATIONAL TOOLS

Each molecular modeling step described above has required the adoption of one or more computational tools. The conformational search (Phase A) is performed in two stages. The first one regards the exploration over all internal degrees of freedom of selector and selectand molecules by Monte Carlo algorithm as implemented in Macromodel v. 4.5.[14] The second stage consists in the minimization of all the structures generated by the MC exploration by using different empirical force fields (E.F.F.). In this stage four E.F.F.s have been considered: MM2*, MM3*, AMBER*,[14] and MMX.[15] Regarding the selectand, the conformational search is performed on one enantiomer only; the structure of the specular isomer (the other enantiomer) is easily obtained by inverting the sign of the *Z* atomic coordinates.

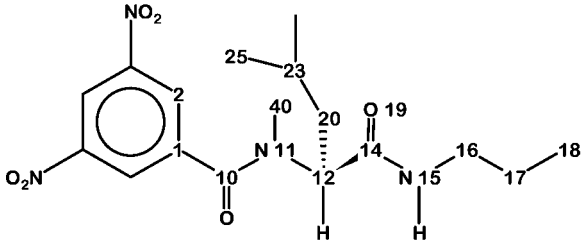
The selection of the conformers to submit to the configurational search is performed by using an energetic criterion. All minima of both selector and selectand within 3RT of the global minimum are selected. The configurational search is performed by applying the previously described "Glob-MolInE" approach. The rigid-body simulations (Phase B) are carried out by the first three modules *L_GRID_6FD*, *SEL_5FD*, and *OPT_6FD*. All minimum configurations obtained are processed by alternate rigid-body and full minimizations (Phase C). The interaction energies computed during the "quasi-flexible" docking are consistent with the empirical force field used to obtain and select the minimized structures of selector and selectand in conformational search.

Statistical thermodynamics are carried out by using the *GM_MTE* module on different statistical ensembles: the first at the *L_GRID_6FD* level, the second at the *OPT_6FD* level, and the last at the *RD_FULL MIN* level.

THE SYSTEM MODELED

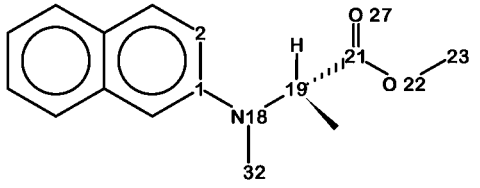
Our simulation was performed referring to a chiral stationary phase (CSP) that has been prepared and studied in detail by Pirkle and coworkers,[16] and is extensively employed today by researchers in organic synthesis and analytical chemistry.

The molecular system we deal with in this article is the complex between (*S*)-N-(3,5-dinitrobenzoyl) leucine *n*-propylamide (*S*)-**1** and the (*S*)-N-(2-naphthyl)-alanine methyl ester (*S*)-**2** (Fig. 3). This system is appropriate to test our methodology because it represents a well-documented example of



(*S*)-N-(3,5-dinitrobenzoyl)-leucine-*n*-propylamide

(*S*)-1



(*S*)-N-(2-naphthyl)alanine methyl ester

(*S*)-2

Definition of rotatable torsions

Torsion angles	Atom number			
τ_{-1}	2	1	10	11
τ_{-2}	1	10	11	12
τ_{-3}	10	11	12	14
τ_{-4}	11	12	14	15
τ_{-5}	12	14	15	16
τ_{-6}	14	15	16	17
τ_{-7}	15	16	17	18
τ_{-8}	11	12	20	23
τ_{-9}	12	20	23	25

Definition of rotatable torsions

Torsion angles	Atom number			
τ_{-10}	2	1	18	19
τ_{-11}	1	18	19	21
τ_{-12}	18	19	21	22

FIGURE 3. Structures of the system modeled and definition of rotatable torsions: Selector (SO), (*S*)-**1**; Selectand (SA), (*S*)-**2**.

enantioselective recognition. (a) The structure of a crystalline 1:1 complex of (*S*)-**1** and (*S*)-**2** has been determined by X-ray crystallography.[16c] Salient features of the complex are two H-bonds and a π-stacking arrangement of the naphthyl and DNB rings along with a loosely H-bonded EtOH molecule also present in the crystal. Intercomplex interactions are not observed in the solid state, thus allowing us to use the solid-state structures for comparison with our gas-phase complexation results. (b) NOE experiments[16a, 16b] highlight other relevant details of the molecular recognition in solution between (*S*)-**1** and (*S*)-**2**; in particular, they give detailed information on the relative orientation of the aromatic rings. (c) The large enantioselectivity factor (α)[16a, 16c] of (*S*)-**1** towards the enantiomers of **2** represents a significant test to highlight the "Glob-MolInE" performance both from a qualitative point of view (prediction of the enantiomers elution orders) and quantitative assessment [thermodynamic parameters evaluation ($\Delta\Delta G$, $\Delta\Delta H$, and $\Delta\Delta S$)]. In fact, many computational studies carried out by other researchers already reported in the literature[7] can be compared with our approach and evaluated on the basis of unquestionable experimental data.[16]

Results

SEARCH FOR THE MOST SUITABLE EMPIRICAL FORCE FIELD AND VALIDATION OF THE "Glob-MolInE" APPROACH

In a preliminary stage the selection of the MMX Force Field is carried out. In our particular case, the coordinates of the X-ray structure are used for both the Empirical Force Field and "Glob-MolInE" validation, because no appreciable interactions between neighboring complexes occur in the solid state. The strategy is articulated according to three main steps; results obtained with the MMX Force Field are reported in Figure 4, the details of which are described below.

1. The first step consists in the evaluation of the Force Field ability to correctly model the experimental complex under consideration; intermolecular orientation (under the control of the nonbonding interactions: van der Waals, H-bonding, and electrostatics) and internal coordinates optimization (Full Force Field) are both taken into account. First, an optimization process on the relative orientation is carried out by the module *OPT_6FD* on the [(*S*)-**1**/(*S*)-**2**]$_{\text{X-Ray}}$ complex considering both (*S*)-**1** and (*S*)-**2** as rigid bodies (path I, center column, Fig. 4). At the end of this process, the geometry of the complex is compared with that of the X-ray complex [RMS = 0.13 Å]. Second, a full minimization is applied to the same X-ray complex (path II, center column, Fig. 4), which has been also submitted to three cycles of rigid docking and Full minimization in an alternate iterative fashion [$\Delta E_i = -23.10$; RMS = 0.29 Å]. The obtained results show that the MMX Force Field describes quite correctly both the intermolecular interactions/orientations (nonbonding interactions are correctly handled) and also the overall molecular geometry of the selector and the selectand in the complex at internal coordinate level. Thus, the MMX Force Field is appropriate for the treatment of our molecular system, at least with regard to the solid state.
2. In the second step (Path **a**, Fig. 4) the structures of (*S*)-**1** and (*S*)-**2** obtained by decomplexation of the X-ray structure [(*S*)-**1**/(*S*)-**2**]$_{\text{X-Ray}}$ have been used to carry out a systematic search of the most representative relative orientations with the rigid docking approach (*L_GRID_6FD* + *SEL_5FD* + *OPT_6FD* modules); obviously, during this search the internal coordinates of both (*S*)-**1** and (*S*)-**2** are not modified. The *L_GRID_6FD* + *SEL_5FD* + *OPT_6FD* modules can be considered efficient if the best orientation obtained during the calculations is similar to or the same as that of the experimental one (X-ray structure). In our case, the most stable structure obtained [Pop. ~99%; $\Delta E_i = -23.25$; RMS = 0.13 Å] by calculation is in very good agreement with the experimental data and is consistent with the energy minimization of the initial X-ray complex.
3. At this point we have validated both the Empirical Force Field and the performance of the rigid docking approach in the generation and selection of the most representative diastereomeric complex. It is now necessary to test the *RD_FULL MIN* module as well, with the aim, in particular, to verify the ability of all the protocol in the evaluation of possible induced fit changes deriving from the complexation process because the isolated molecule's conformations can be very different from those found in the complex. Thus, a preliminary separation (decomplexation) (Path **b**, Fig. 4),

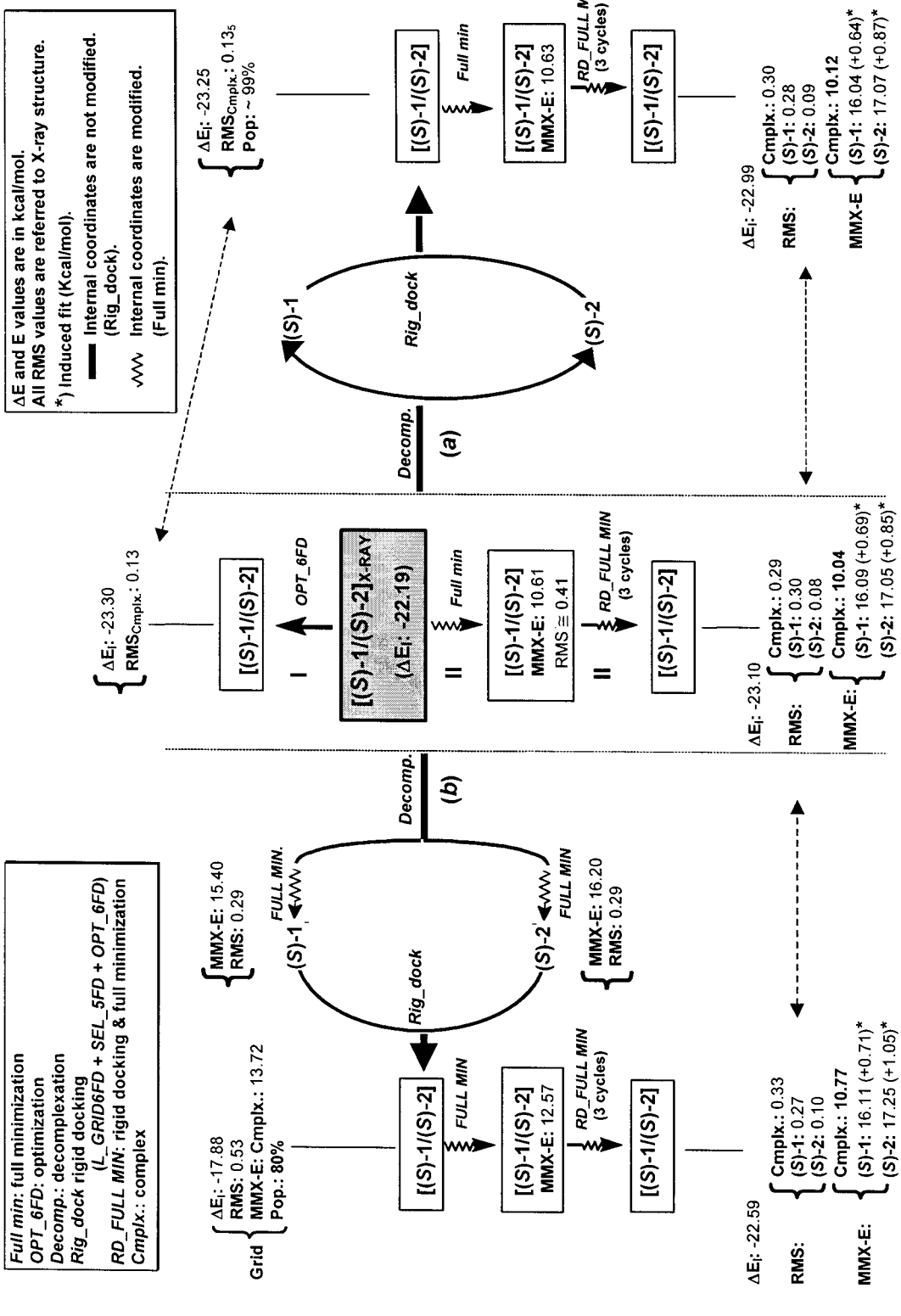


FIGURE 4. Schematic representation of the search for the most suitable Empirical Force Field (E.F.F.) and of "Glob-MollnE" validation (see text for details).

followed by a full minimization on the separated molecules, has been carried out. In this step, the minimized structures are the starting point for the rigid docking.

Now we find the most representative (Pop. 80%) complex so obtained to differ substantially from the experimental one [$\Delta E_i = -17.88$; RMS = 0.53 Å] because at this stage the calculation does not consider variations in the internal coordinates of the molecules that have been previously subjected to relaxation of the internal coordinates. In fact, the interaction energy and the RMS values are fairly high. This complex is then subjected to three Full minimization and Rigid docking sequential cycles. At the end of this process (*RD_FULL MIN)* the interaction energy ($\Delta E_i = -22.59$) and the RMS (0.33 Å) fall notably, and the final structure is now very similar to the experimental one. That is to say, for complexes with a large interaction energy, the minimum energy conformations of the isolated molecules are expected to change when complexation takes place.

Thus, the complete proposed protocol of calculation allows us to validate the force field both for the intermolecular interaction/orientation and for the internal coordinates optimization and thus to demonstrate that the rigid docking approach is able to reproduce, with remarkable qualitative and quantitative exactness, the X-ray geometry orientation. Moreover the *RD_FULL MIN* module, applied after the rigid docking is able to generate and evaluate possible induced fit changes that take place upon coordination of flexible selectors with flexible selectands.

The Reciprocity Concept: Computational Point of View

Simply stated, the reciprocity concept means that if one enantiomer of a given molecule can distinguish between the enantiomer of a second, then one enantiomer of the latter should be able to distinguish between the enantiomer of the former.

From the computational point of view, this requires that the final results must be identical when selector and selectand are interchanged; so, following this concept we have considered (*S*)-**2** as the stationary molecule with its center of mass in the origin of a system of polar coordinates, while (*S*)-**1** has been made to roll on the van der Waals surface of (*S*)-**2** according to the previously described protocol. The results, in terms of geometry, interaction energy and RMS deviation from for the experimental results, are coincident with those obtained keeping (*S*)-**1** static and (*S*)-**2** as the moving molecule. The final results do not depend on the original starting orientations of the two docking molecules, and they can be interconverted according to the reciprocity concept.

Docking between the Pirkle's Complex [(*S*)-**1**/(*S*)-**2**]$_{\text{X-Ray}}$ and Ethanol

Another accurate test for the "Glob-MolInE" approach is carried out taking into account the position of the ethanol molecule described in the X-ray structure of Pirkle's complex[16c] to study the interactions by computational methods. In this test the populated conformations of the alcohol molecule (within 3 kcal/mol of its global minimum) are docked with the [(*S*)-**1**/(*S*)-**2**]$_{\text{X-Ray}}$ complex using both the rigid (Phase B) and the full relaxation modules (Phase C) of "Glob-MolInE"; the final statistical thermodynamic evaluation module (*GM_MTE*) gives 18 configurations for the selected complex within 2.0 kcal/mol of the global minimum, representing more than 90% of the total population. This set of configurations is selected and analyzed to study the location of the ethanol molecule within the complex and to monitor the conformational changes of (*S*)-**1** and (*S*)-**2** induced by that solvent molecule. We analyzed the results essentially in terms of H-bonding contribution, considering this interaction as the main driving force of the solvation process. All the possible H-bonding acceptor/donor interaction sites in the solvent and in the complexed molecules are classified: a total of nine hydrogen-bonding sites [(*S*)-**1**$_{\text{HB-A}_{1-4}}$, (*S*)-**2**$_{\text{HB-A}_{1-2}}$, (*S*)-**1**$_{\text{HB-D}_{1-2}}$, (*S*)-**2**$_{\text{HB-D}_1}$] in the [(*S*)-**1**/(*S*)-**2**]$_{\text{X-Ray}}$ complex and two sites on the ethanol molecule are easily identified (Fig. 5). All possible interactions, both inter- and intramolecular, deriving from the potential H-bonding sites of the studied molecular system [(*S*)-**1**, (*S*)-**2**, EtOH] are analyzed (threshold = 2.5 Å) for each of the 18 most stable configurations selected previously. The obtained results show only nine different acceptor/donor relevant interactions; the graphical representation in the form of a 9 × 18 matrix is shown in Figure 6.

Based on the specific sites involved in the interactions, four different H-bonding patterns were identified (HB_1,4) (Figs. 6 and 7).

In the first pattern (HB_1, Fig. 7a) the EtOH molecule establishes two H-bonds with the dinitrobenzoyl carbonyl oxygen and the terminal amidic NH

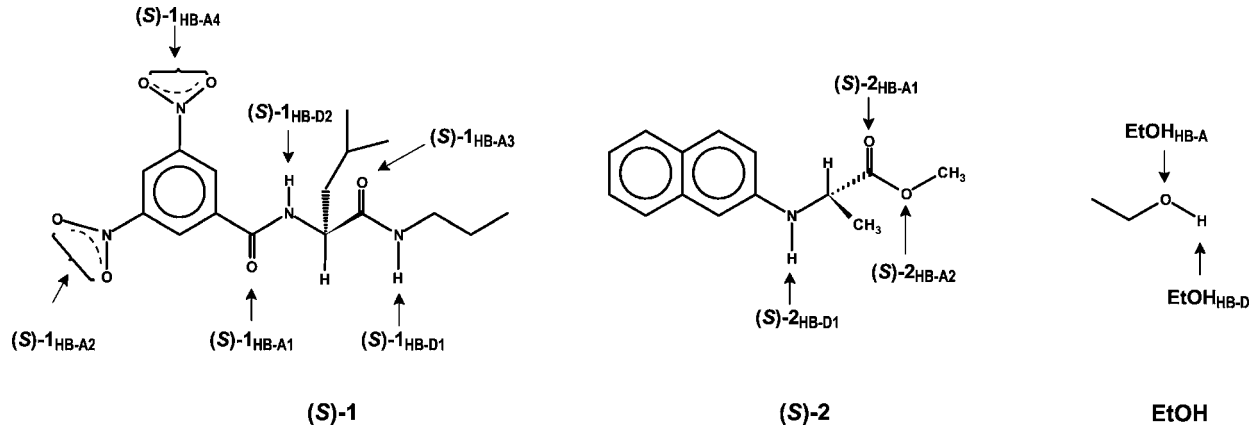


FIGURE 5. Hydrogen-bond atom classification: labels $_A$ and $_D$ mean, respectively, acceptor and donor.

of (*S*)-**1**. Interestingly, the intramolecular H-bond (*S*)-$\mathbf{1}_{D1}$/(*S*)-$\mathbf{1}_{A1}$ is disrupted in this configuration. The Boltzmann population at room temperature of this pattern amounts to 52% and is found in the 1st, 2nd, 5th, 9th, and 10th minimum energy configurations.

The second pattern (HB_2, Fig. 7b) is found in ten different configurations (3rd, 4th, 7th, 11th, 12th, 13th, 14th, 15th, 17th, and 18th; Boltzmann population equal to 33%): here, the EtOH molecule is engaged in a single H-bond with the oxygen of a nitro group of (*S*)-**1**, while the H-bond network holding together the complex remains unaffected.

The third pattern (HB_3, Fig. 7c) is found only in the sixth and eighth configurations (7% Boltzmann population) and features two H-bonds between the alcohol and the NH of (*S*)-**2** and the C-terminal carbonyl oxygen of (*S*)-**1**. The intermolecular H-bond (*S*)-$\mathbf{2}_{D1}$/(*S*)-$\mathbf{1}_{A3}$ is absent in this configuration.

In the fourth pattern (HB_4, Fig. 7d) found in the 16th configuration (Boltzmann population equal to 1%) a single H-bond is formed by the EtOH molecule with a nitro group oxygen, while the other H-bonding interactions between (*S*)-**1** and (*S*)-**2** are not altered.

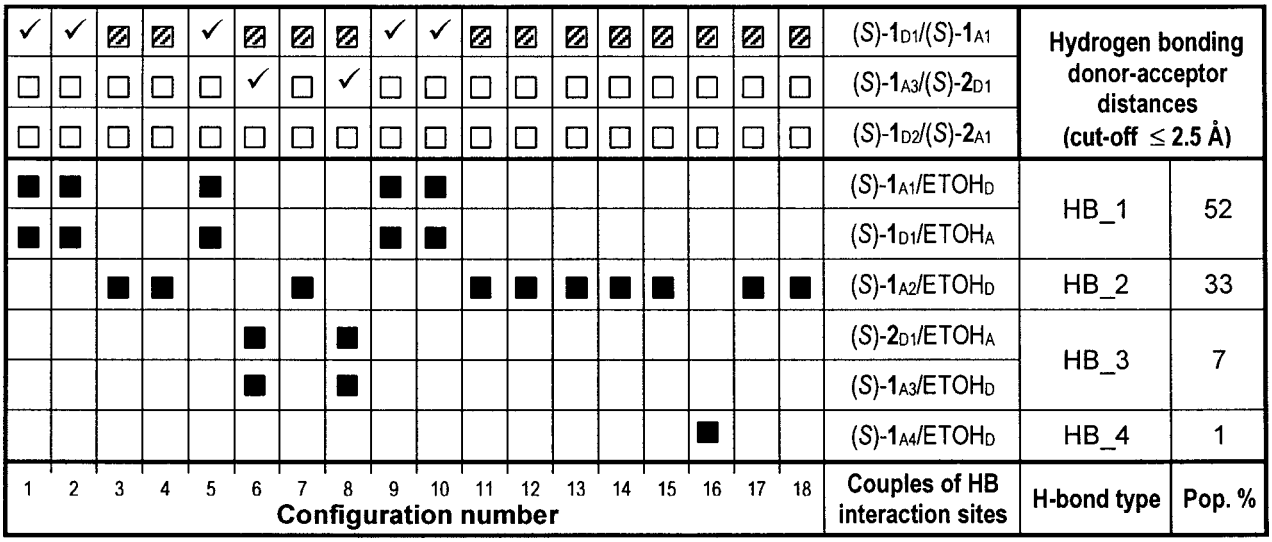


FIGURE 6. Hydrogen-bonding plot analysis: filled squares represent hydrogen bond interactions between EtOH and the [(*S*)-**1**/(*S*)-**2**] complex; diagonal filled squares represent the intramolecular hydrogen bond interactions in (*S*)-**1**; empty squares represent the intermolecular hydrogen bond interactions between (*S*)-**1** and (*S*)-**2** in the [(*S*)-**1**/(*S*)-**2**]·EtOH complex; the symbol ✓ means the absence of a hydrogen bond in the [(*S*)-**1**/(*S*)-**2**] complex due to the solvation.

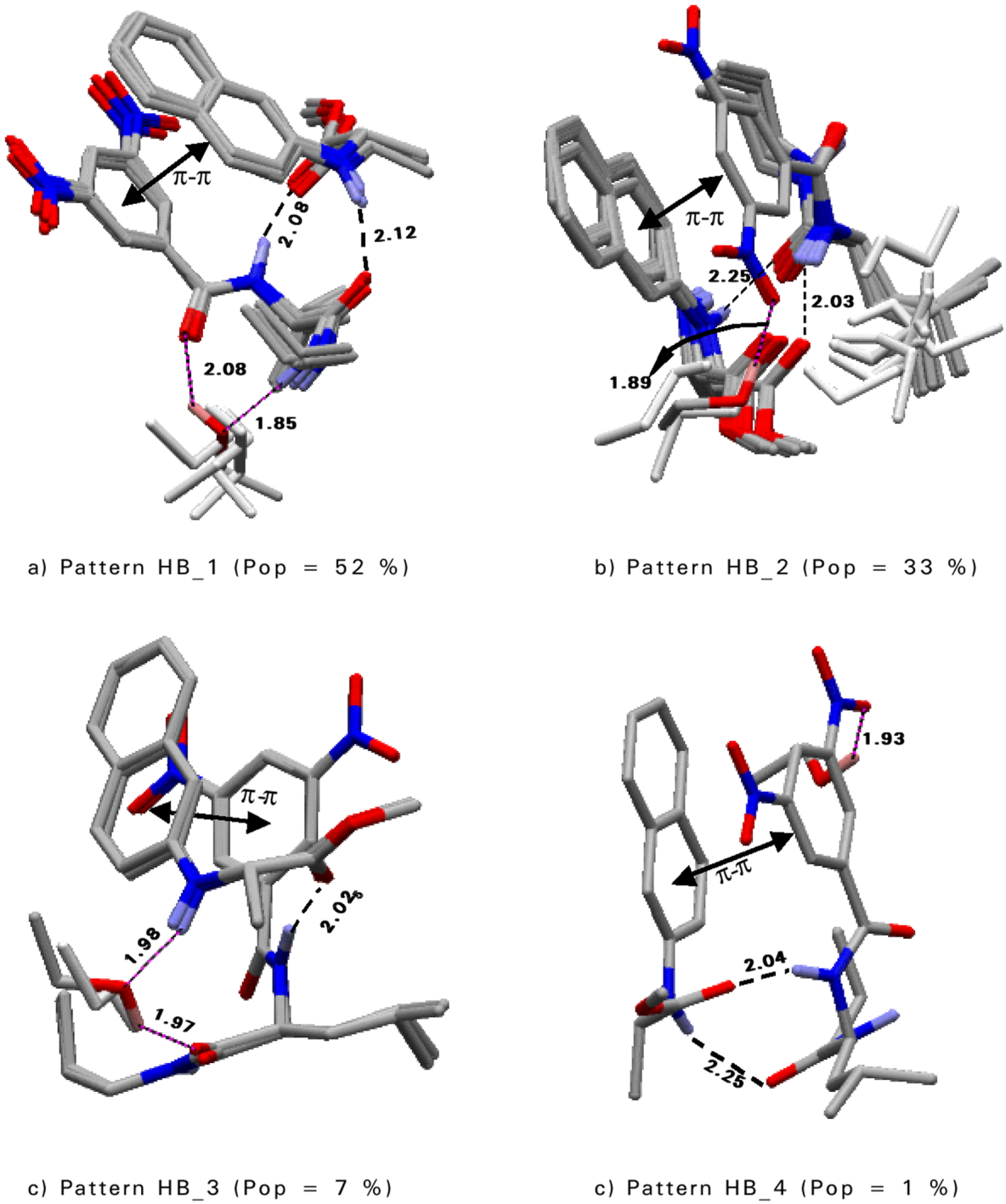


FIGURE 7. Polytube superimposition of the four patterns of H-bonding in the "Glob-MolInE" docking of EtOH and the [(*S*)-**1**/(*S*)-**2**] complex. Each relative minimum configuration of the alcohol is maintained in standard CPK colors and the other orientations are depicted in white for clarity. (a) pattern HB_1; (b) pattern HB_2; (c) pattern HB_3; (d) pattern HB_4.

Thus, more than 50% of the obtained configurations can be explained in terms of HB_1, in which the EtOH molecule is located in the same position found experimentally in the solid state,[16c] loosely hydrogen-bonded to the dinitrobenzoyl carbonyl oxygen of (*S*)-**1**. The other three HB patterns locate the alcohol molecule close to H-bond donor/acceptor sites within the complex, while

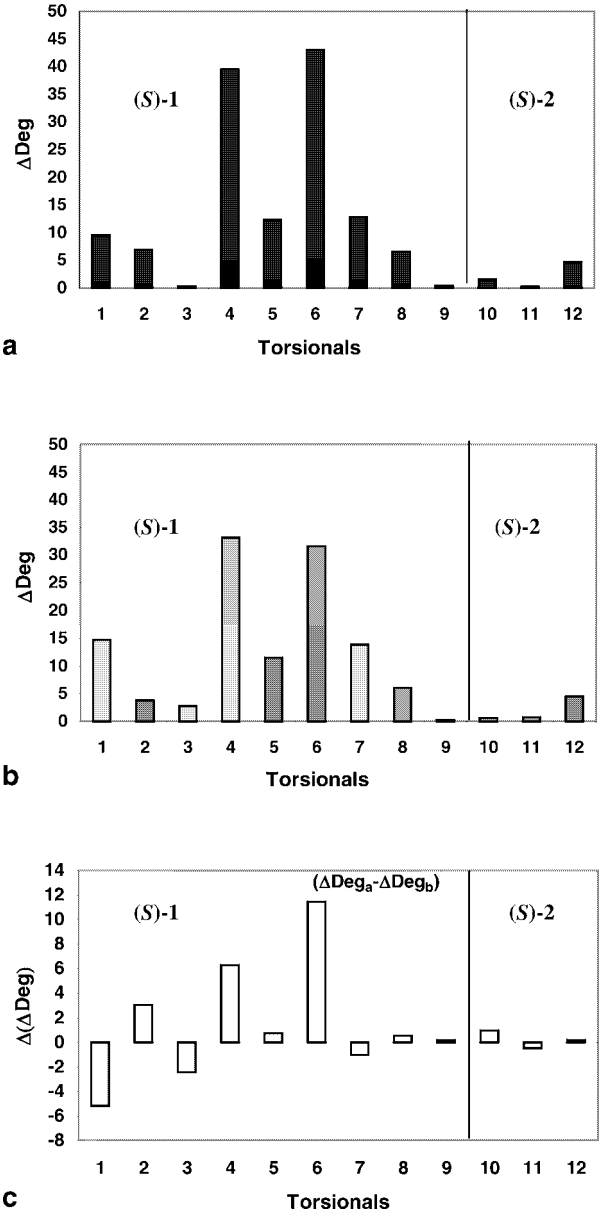


FIGURE 8. (a) Torsional angles variations between $[(S)\text{-}\mathbf{1}/(S)\text{-}\mathbf{2}]_{\text{X-Ray}}$ and $[(S)\text{-}\mathbf{1}/(S)\text{-}\mathbf{2}]_{\text{X-Ray(min)}}$; $\Delta\text{Deg.} = |\tau_{i,\text{X-Ray}} - \tau_{i,\text{X-Ray(min)}}|$; $1 \leq i \leq 12$. (b) Torsional angle variations between $[(S)\mathbf{1}/(S)\text{-}\mathbf{2}]_{\text{X-Ray}}$ and $\{[(S)\text{-}\mathbf{1}/(S)\text{-}\mathbf{2}]\cdot\text{EtOH}\}_{\text{avgd.,18}}$; $\Delta\text{Deg.} = |\tau_{i,\text{X-Ray}} - \tau_{i,\{[(S)\text{-}\mathbf{1}/(S)\text{-}\mathbf{2}]\cdot\text{EtOH}\}_{\text{avgd.,18}}}|$; $1 \leq i \leq 12$. (c) Difference between (a) and (b); $\Delta(\Delta\text{Deg.}) = \Delta\text{Deg.}$ (a) $-\ \Delta\text{Deg.}$ (b). Positive values indicate a better fit with the X-ray complex.

other configurations (7% Boltzmann population) are found above the 2 kcal threshold. This result thus explains the presence of one disordered ethanol molecule per molecule of complex, as found experimentally.

Finally, we also evaluated the conformational variations induced by the ethanol on the selector and selectand molecules in the final fully relaxed solvated complex. To do this, we first compared the torsional values of the $[(S)\text{-}\mathbf{1}/(S)\text{-}\mathbf{2}]_{\text{X-Ray}}$ complex ($\tau_{i,\text{X-Ray}}$; $1 \leq i \leq 12$) to the corresponding values in the same complex after geometry optimization in the absence of the ethanol molecule ($\tau_{i,\text{X-Ray(min)}}$).

The results (Fig. 8a) point out that the energetic minimization changes appreciably only eight torsion angles (ns. 1, 2, 4, 5, 6, 7, 8, 12) out of the 12 chosen, causing a distortion in the coordinates characterized by 0.41 RMS. This distortion is essentially present in (*S*)-**1**.

At the same time we also calculated the same 12 averaged torsional values, weighted on the Boltzmann population at room temperature (298°K) for the 18 most stable selected configurations ($\tau_{i,\{[(S)\text{-}\mathbf{1}/(S)\text{-}\mathbf{2}/]\cdot\text{EtOH}\}_{\text{avgd.,18}}}$) and compared them with the corresponding torsional value $\tau_{i,\text{X-Ray}}$ (Fig. 8b).

In this case, the inclusion of the ethanol molecule in the docking procedure, after full minimization, induces a reduction (Fig. 8b and c) both in the angular deviations for most of the torsionals ($i = 2, 4, 5, 6, 8, 9, 10$), and in the RMS value (0.20) (Fig. 9), confirming that the procedure of "quasi-flexible" docking can keep track of induced fit effects of specific nonbonded interactions at the intermolecular level. In particular, we find that the H-bonded EtOH molecule contained in the crystal stabilizes the $[(S)\mathbf{1}/(S)\text{-}\mathbf{2}]$ complex. As the alcohol molecule interacts with the complex, some torsional angles of $(S)\mathbf{1}$ (Fig. 8c) are slightly changed, thus facilitating the H-bonding interactions between selector and selectand.

So far, applications of the above protocol are being carried out at present for several relevant flexible chiral molecular selectors and selectands containing many torsional degrees of freedom; the preliminary results are very interesting, and are in agreement both with spectroscopically derived recognition mechanism and with thermodynamic quantities extracted from chromatographic data.

Acknowledgments

The authors thank Prof. W. H. Pirkle for providing X-ray coordinates of the complex, Prof. W. C. Still for making MacroModel available to them, and Prof. K. Lipkowitz for the helful discussion during the preparation of the article.

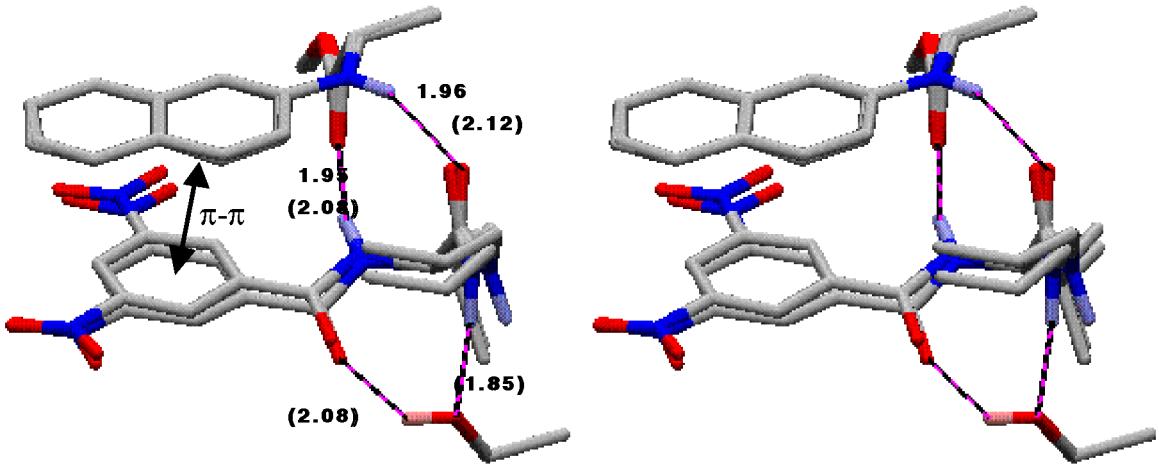


FIGURE 9. Comparison between [(*S*)-**1**/(*S*)-**2**]$_{X\text{-}Ray}$ and the global minimum of the [(*S*)-**1**/(*S*)-**2**]·EtOH solvated complex. Polytube model stereoview.

References

1. (a) Souter, R. W. Chromatographic Separations of Stereoisomers; CRC Press: Boca Raton, 1985; (b) Zief, M.; Crane, L. J. Chromatographic Chiral Separations (Chromatographic Science Series, Vol. 40); Marcel Dekker: New York, 1988; (c) Lough, W. J. Chiral Liquid Chromatography; Blackie & Sons: Glasgow, 1989; (d) Krstulovic, A. M. Chiral Separations by HPLC: Application to Pharmaceutical Compounds, Ellis Horwood: Chichester, 1989; (e) Stevenson, D.; Wilson, I. D. Recent Advances in Chiral Separations, Plenum: New York, 1990; (f) Allenmark, S. G. Chromatographic Enantioseparations: Methods and Applications, Ellis Horwood: Chichester, 1991; (g) Ahuja, S. Chiral Separations by Liquid Chromatography, ACS Symposium Series: Washington, DC, 1991; (h) Subramanian, G. A Practical Approach to Chiral Separations by Liquid Chromatography, VCH: New York, 1994; (i) Lindner, W. Chiral Operations. Fundamental Aspects and Applications, J Chromatogr A 1994, 666; (j) Pirkle, W. H.; Pochapsky, T. C. In Advances in Chromatography; Giddings, J. C.; Grushka, E.; Brown, P. R., Eds.; Marcel Dekker: New York, 1987, Vol. 27, p. 73; (k) Pirkle, W. H.; Pochapsky, T. C. Chem Rev 1989, 89, 347; (l) Taylor, D. R.; Maher, K. J. Chromatogr Sci 1992, 30, 67.
2. (a) Pirkle, W. H.; Liu, Y. J Org Chem 1994, 59, 6911; (b) Welch, C. J. J Chromatogr A 1994, 666, 3; Pirkle, W. H.; Murray, P. G. J Chromatogr 1993, 641, 11; (d) Pirkle, W. H.; Welch, C. J.; Lamm, B. J Org Chem 1992, 57, 3854; (e) Pirkle, W. H.; Däppen, R. J Chromatogr A 1987, 404, 107; Pirkle, W. H; Pochapsky, T. C. J Am Chem Soc 1986, 108, 352; Pirkle, W. H.; House, D. W.; Finn, J. M. J Chromatogr A 1980, 192, 143.
3. (a) Gschwend, D. A.; Kuntz, I. D. J Comput Aided Mol Design 1996, 10, 123; (b) Oshiro, C. M.; Kuntz, I. D.; Scott Dixon, J. J Comput Aided Mol Design 1995, 9, 113.
4. (a) Lipkowitz, K. B.; Raghothama, S.; Yang, J. J Am Chem Soc 1992, 114, 1554; (b) Lipkowitz, K. B.; Baker, B. Anal Chem 1990, 62, 770; (c) Lipkowitz, K. B.; Baker, B.; Zagarra,R. J. Comput Chem 1989, 10, 718; (d) Lipkowitz, K. B.; Zagarra, R. J. J Comput Chem 1989, 10, 595; (e) Lipkowitz, K. B.; Antell, S.; Baker, B. J Org Chem 1989, 54, 5449; (f) Lipkowitz, K. B.; Baker, B.; Larter, R. J Am Chem Soc 1989, 111, 7750; (g) Lipkowitz, K. B.; Demeter, D. A.; Zagarra, R. J.; Larter, R.; Darden, T. J Am Chem Soc 1988, 110, 3446; (h) Lipkowitz, K. B.; Demeter, D. A.; Parish, C. A.; Darden, T. Anal Chem 1987, 59, 1731; (i) Lipkowitz, K. B.; Landwer, J. M.; Darden, T. Anal Chem 1986, 58, 1611.
5. Däppen, R.; Karfunkel, H. R.; Leusen, F. J. J. J Comput Chem 1990, 11, 181.
6. (a) Still, M. G.; Rogers, L. B. Talanta 1990, 37, 599; (b) Still, M. G.; Rogers, L. B. J Comput Chem 1990, 11, 242; (c) Still, M. G.; Rogers, L. B. Talanta 1989, 36, 35.
7. (a) Topiol, S.; Sabio, M. Enantiomer 1996, 1, 251; (b) Sabio, M.; Topiol, S. Chirality 1991, 3, 56; (c) Sabio, M.; Topiol, S. J Am Chem Soc 1989, 111, 4109; (d) Topiol, S. Chirality 1989, 1, 69; (e) Topiol, S.; Sabio, M. J Chromatogr 1989, 461, 129; (f) Sabio, M.; Topiol, S. Int J Quantum Chem 1989, 36, 323; (g) Topiol, S.; Sabio, M.; Moroz, J.; Caldwell, W. B. J Am Chem Soc 1988, 110, 8367.
8. Aerts, J. J Comput Chem 1995, 16, 914.
9. Norinder, V.; Sundholm, E. G. J Liq Chromatogr 1987, 10, 2825.
10. Edge, A. M.; Heaton, D. M.; Bartle, K. D.; Clifford, A. A.; Myers, P. A. Chromatographia 1995, 41, 161.
11. Configuration is defined as any unique orientation of selectand with respect to the selector (in our particular case, it represents a diastereomeric associate at supramolecular level).
12. (a) Kuhl, F. S.; Crippen, G. M.; Friesen, D. K. J Comput Chem 1994, 5, 24; (b) Leach, A. In Reviews in Computational Chemistry; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH: New York, 1991, Vol. 2.
13. Press, W. H.; Teukolsky, S. A.; Wetterling, W. T.; Flannery, B. P. Numerical Recipes in FORTRAN. The Art of Scientific Computing; Cambridge University Press: Cambridge, MA, 1992.
14. Still, C. W. Macromodel, V. 4.5; Columbia University: New York. See also: Mohamadi, F.; Richards, N. G. J.; Guida,

W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. J Comput Chem 1990, 11, 440.

15. We made use of the MMX force field as implemented in the *PC Model* program (Serena Software: Bloomington, IN). (See also: Gajewski, J. J.; Gilbert, K. K.; McKelvey, J. Advances in Molecular Modelling; JAI Press: Greenwich, CT, 1992, Vol. 2.) The MMX adopted has two modifications from the original empirical force field. The first regards the nitro group. The van der Waals radii of both oxygen atoms (type 7 and 42) are set at 1.74 Å and the dipole moment values between those atoms and the nitrogen (type 41) are set at 2.35 D. This parametrization has been performed modifying the parameter database MMX.STO. The second modification regard the sp^2-sp^2 treatment in the van der Waals energy term computation. This is turned on in all the calculations.
16. (a) Pirkle, W. H.; Pochapsky, T. C. J Am Chem Soc 1986, 108, 5627; (b) Pirkle, W. H.; Pochapsky, T. C. J Am Chem Soc 1987, 109, 5975; (c) Pirkle, W. H.; Burke, J. A., III; Wilson, S. R. J Am Chem Soc 1989, 111, 9222.